



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2016

iCOBRA: open, reproducible, standardized and live method benchmarking

Soneson, Charlotte ; Robinson, Mark D

DOI: <https://doi.org/10.1038/nmeth.3805>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-124743>

Journal Article

Accepted Version

Originally published at:

Soneson, Charlotte; Robinson, Mark D (2016). iCOBRA: open, reproducible, standardized and live method benchmarking. *Nature Methods*, 13(4):283.

DOI: <https://doi.org/10.1038/nmeth.3805>

iCOBRA: open, reproducible, standardized and live method benchmarking

Charlotte Sonesson^{1,2} and Mark D. Robinson^{1,2,*}

¹Institute of Molecular Life Sciences, University of Zurich, Switzerland

²SIB Swiss Institute of Bioinformatics, Zurich, Switzerland

*Corresponding author, mark.robinson@imls.uzh.ch

Modern life science research tasks often involve ranking or classifying items. For example, differential expression studies can rank genes by the estimated p-value or, using a cutoff on the p-values, classify genes as either “significantly different” or “not significantly different” between conditions of interest. A wide range of computational methods dedicated to these tasks exist^{1–3}, many of them relying on accurate quantifications of underlying entities such as abundance levels. As methods are developed and refined, static benchmarking studies quickly become outdated. Moreover, a standard way to present results from method comparisons is lacking and raw data are not always made available. This often makes it difficult for method researchers to reproduce published evaluations or explore them from different angles. Here, we present iCOBRA (interactive COmparative evaluation of Binary classification and RAnking methods), a benchmarking platform for both users and developers of methods, promoting open, standardized and reproducible evaluations. iCOBRA consists of an R package as well as a flexible, interactive web application that can rapidly evaluate methods for binary classification, ranking and continuous target estimation against a ground truth. In addition, we have collected a set of benchmarking datasets in standard formats (a link is provided on <https://github.com/markrobinsonuzh/iCOBRA>), to lower barriers for new method developers but also to facilitate standardized method evaluations in the future. We envision this resource to be extended over time and encourage the community to contribute their data (e.g., simulations) and method assessments. In Supplementary Note 1, we show how iCOBRA can be used to exactly reproduce and visualize results from recent benchmarking studies.

iCOBRA's web application (Fig. 1) is based on the Shiny framework, and can be run via our public server (accessible from <https://github.com/markrobinsonuzh/iCOBRA>), which makes it platform-agnostic and eliminates the need for knowledge about installing or running R. Extensive documentation is included in the app and is reproduced in Supplementary Note 2. Underlying the application is an R package (available via Bioconductor) that can be used both to run the interactive application locally and to generate result visualizations directly from the R console, facilitating both interactive exploration and integration within programming pipelines. In contrast to R packages dedicated to evaluating classifiers (e.g., ROCR⁴), which generate static performance plots, the interactivity provided by the Shiny framework lets the user include or exclude methods from a comparison, change the appearance of the plots or stratify the results by a provided annotation with minimal effort. The input format is simple and generic (tab-delimited text files), leading to an increased ease and range of use compared to other performance evaluators (e.g., compcodeR⁵), whose data representation format and/or choice of evaluation metrics are specifically tailored to certain types of data. The application accepts several input types (nominal p-values, adjusted p-values and a general “score”), allowing for flexibility beyond existing applications like BDTcomparator⁶, which compares two categorizations and is thus strictly limited to classification evaluation.

Competing financial interests

The authors declare that they have no competing financial interests.

References

1. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
2. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10**, 1200–1202 (2013).
3. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-

- 61 seq data. *Genome Res.* **22**, 2008–2017 (2012).
- 62 4. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: Visualizing classifier
63 performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
- 64 5. Sonesson, C. compcodeR - an R package for benchmarking differential expression
65 methods for RNA-seq data. *Bioinformatics* **30**, 2517–2518 (2014).
- 66 6. Fijorek, K., Fijorek, D., Wisniowska, B. & Polak, S. BDTcomparator: A program for
67 comparing binary classifiers. *Bioinformatics* **27**, 3439–3440 (2011).

68

69 Figure Legends

70 **Figure 1. Screenshot of the iCOBRA interactive application interface.** The left
71 sidebar contains the input controls, while the tabs in the main panel display different
72 aspects of the performance evaluation.
73

- Truth
- Results

Add file with results

Choose Filemerged_resul...sophila.txt

Upload complete

Select methods

- ☐ cuffdiff_ensembl
- ☒ dexseq_casper
- ☒ dexseq_featurecounts_flat
- ☐ dexseq_featurecounts_noflat
- ☒ dexseq_htseq_nomerge
- ☐ dexseq_htseq
- ☒ dexseq_kallisto
- ☒ dexseq_miso_assignable
- ☒ dexseq_splicinggraph
- ☒ dexseq_tophat_junc
- ☐ rMATS_junction
- ☐ Calculate performance based only on features shared between truth and result tables.

✈ Start calculation!

- Plot settings

Select color palette

Set1 (max 8 methods)

Display full curve and/or points in FDR/TPR and FDR/NBR plots

- ☐ curve
- ☒ points

Split plots into panels by stratifying variable

- ☒ split

FDR thresholds

